07 624120

11509-28

# Photolabile Nucleoside and Peptide Protecting Groups and Improved Polymer Synthesis

~~VERY LARGE SCALE IMMOBILIZED POLYMER SYNTHESIS~~

## CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of ~~pending~~ application Serial No. 492,462, filed March 7, 1990, which is a continuation-in-part of pending application Serial No. 362,901, filed June 7, 1989, assigned to the assignee of the present application, and incorporated herein by reference for all purposes. This application is related to applications with Serial Nos. ~~07/626,730~~ and ~~07/624,114~~ 642,114, ~~(Attorney Docket Nos. 11509-26 and 11509-30)~~, filed on the same day as the present application, and also incorporated herein by reference for all purposes. Microfiche Appendices A and B are attached, including 3 sheets of microfiche comprising 226 frames.



## BACKGROUND OF THE INVENTION

The present invention relates to the field of polymer synthesis. More specifically, the invention provides a reactor system, a masking strategy, photoremovable protective groups, data collection and processing techniques, and applications for light directed synthesis of diverse polymer sequences on substrates.

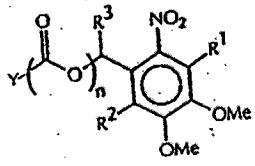
## SUMMARY OF THE INVENTION

Methods, apparatus, and compositions for synthesis and use of ~~techniques for~~ diverse polymer sequences on a substrate are disclosed, as well as applications thereof.

According to one aspect of the invention, an improved reactor system for synthesis of diverse polymer sequences on a substrate is provided. According to this embodiment the invention provides for a reactor for contacting reaction fluids to a substrate; a system for delivering selected reaction fluids to the reactor; a translation stage for moving a mask or substrate from at least a first relative location relative to a second relative location; a light for illuminating the substrate through a mask at selected times; and an appropriately programmed digital computer for selectively directing a flow of fluids from the reactor system, selectively activating the translation stage, and selectively illuminating the substrate so as to form a plurality of diverse polymer sequences on the substrate at predetermined locations.

The invention also provides a technique for selection of linker molecules in ~~VLSIPS~~. According to this aspect of the invention, the invention provides a method of screening a plurality of linker polymers for use in binding affinity studies. The invention includes the steps of forming a plurality of linker polymers on a substrate in selected regions; the linker polymers formed by the steps of recursively: on a surface of a substrate, irradiating a portion of the selected regions to remove a protective group, and contacting the surface with a monomer; contacting the plurality of linker polymers with a ligand; and contacting the ligand with a labeled receptor.

According to another aspect of the invention, improved photoremovable protective groups are provided. According to this aspect of the invention a compound having the formula:



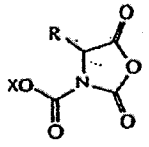
wherein n = 0 or 1; Y is selected from the group consisting of an oxygen of the carboxyl group of a natural or unnatural amino acid, an amino group of a natural or unnatural amino acid, or the C-5' oxygen group of a natural or unnatural deoxyribonucleic or ribonucleic acid; $R^1$ and $R^2$ independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxyl, thiol, thioether, amino, nitro, carboxyl, formate, formamido, sulfido, or phosphido group; and $R^3$ is a alkoxy, alkyl, aryl, hydrogen, or alkenyl group is provided.

The invention also provides improved masking techniques for the VLSIPS™ methodology. According to one aspect of the masking technique, the invention provides an ordered method for forming a plurality of polymer sequences by sequential addition of reagents comprising the step of serially protecting and deprotecting portions of the plurality of polymer sequences for addition of other portions of the polymer sequences using a binary synthesis strategy.

Improved data collection equipment and techniques are also provided. According to one embodiment, the instrumentation provides a system for determining affinity of a receptor to a ligand comprising: means for applying light to a surface of a substrate, the substrate comprising a plurality of ligands at predetermined locations, the means for

~~applying directing light~~ providing simultaneous illumination at a plurality of the predetermined locations; and an array of detectors for detecting light fluoresced at the plurality of predetermined locations. The invention further provides for improved data analysis techniques including the steps of exposing fluorescently labelled receptors to a substrate, the substrate comprising a plurality of ligands in regions at known locations; at a plurality of data collection points within each of the regions, determining an amount of light fluoresced from the data collection points; removing the data collection points deviating from a predetermined statistical distribution; and determining a relative binding affinity of the receptor to remaining data collection points.

Protected amino acid N-carboxy anhydrides for use in polymer synthesis are also disclosed. According to this aspect ~~of~~ the invention provides a compound having the formula:




where R is a side chain of a natural or unnatural amino acid and X is a photoremovable protecting group.

A further understanding of the nature and advantages of the inventions herein may be realized by reference to the remaining portions of the specification and the attached drawings.

## BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 schematically illustrates light-directed spatially-addressable parallel chemical synthesis;

Fig. 2 schematically illustrates one example of light-directed peptide synthesis;

Fig. 3 is a three-dimensional representation of a portion of the checkerboard array of YGGFL and PGGFL;

Fig. 4 schematically illustrates ~~the software for the~~ automated system for synthesizing diverse polymer sequences;

Fig. 5a and 5b illustrate operation of a program for polymer sythesis;

Fig. 6 is a schematic illustration of a "pure" binary masking strategy;

Fig. 7 is a schematic illustration of a gray code binary masking strategy;

Fig. 8 is a schematic illustration of a modified gray code binary masking strategy;

Fig. 9a schematically illustrates a masking scheme for a four step synthesis;

Fig. 9b schematically illustrates synthesis of all 400 peptide dimers;

Fig. 10 is a coordinate map for the ten-step binary synthesis;

Fig. 11 schematically illustrates a data collection system;

Fig. 12 is a block diagram illustrating the architecture of the data collection system;

Fig. 13 is a flow chart illustrating operation of software for the data collection/analysis system; and

Fig. 14 illustrates a three-dimensional plot of intensity versus position for light directed synthesis of a dinucleotide.

## DESCRIPTION OF THE PREFERRED EMBODIMENTS

CONTENTS

I. Definitions

Certain terms used herein are intended to have the following general definitions:

1. Complementary: Refers to the topological compatibility or matching together of interacting surfaces of a ligand molecule and its receptor. Thus, the receptor and its ligand can be described as complementary, and furthermore, the contact surface characteristics are complementary to each other.

2. Epitope: The portion of an antigen molecule which is delineated by the area of interaction with the subclass of receptors known as antibodies.

3. Ligand: A ligand is a molecule that is recognized by a particular receptor. Examples of ligands that can be investigated by this invention include, but are not restricted to, agonists and antagonists for cell membrane receptors, toxins and venoms, viral epitopes, ~~hormones (e.g., opiates, steroids, etc.),~~ hormones, hormone receptors, peptides, enzymes, enzyme substrates, cofactors, ~~drugs,~~ lectins, sugars, oligonucleotides, nucleic acids, oligosaccharides, proteins, and monoclonal antibodies.

4. Monomer: A member of the set of small molecules which can be joined together to form a polymer. The set of monomers includes but is not restricted to, for example, the set of common L-amino acids, the set of D-amino acids, the set of synthetic amino acids, the set of nucleotides and the set of pentoses and hexoses. As used herein, monomers refers to any member of a basis set for synthesis of a polymer. For example, dimers of the 20 naturally occurring L-amino acids form a basis set of 400

monomers for synthesis of polypeptides. Different basis sets of monomers may be used at successive steps in the synthesis of a polymer. Furthermore, each of the sets may include protected members which are modified after synthesis.

5. Peptide: A polymer in which the monomers are alpha amino acids and which are joined together through amide bonds and alternatively referred to as a polypeptide. In the context of this specification it should be appreciated that the amino acids may be the L-optical isomer or the D-optical isomer. Peptides are often two or more amino acid monomers long, and often more than 20 amino acid monomers long. Standard abbreviations for amino acids are used (e.g., P for proline). These abbreviations are included in Stryer, Biochemistry, Third Ed., 1988, which is incorporated herein by reference for all purposes.

6. Radiation: Energy which may be selectively applied including energy having a wavelength of between $10^{-14}$ and $10^{4}$ meters including, for example, electron beam radiation, gamma radiation, x-ray radiation, ultraviolet radiation, visible light, infrared radiation, microwave radiation, and radio waves. "Irradiation" refers to the application of radiation to a surface.

7. Receptor: A molecule that has an affinity for a given ligand. Receptors may be naturally-occurring or manmade molecules. Also, they can be employed in their unaltered state or as aggregates with other species. Receptors may be attached, covalently or noncovalently, to a binding member, either directly or via a specific binding substance. Examples of receptors which can be employed by this invention include, but are not restricted to, antibodies,

cell membrane receptors, monoclonal antibodies and antisera reactive with specific antigenic determinants (such as on viruses, cells or other materials), drugs, polynucleotides, nucleic acids, peptides, cofactors, lectins, sugars, polysaccharides, cells, cellular membranes, and organelles. Receptors are sometimes referred to in the art as anti-ligands. As the term receptors is used herein, no difference in meaning is intended. A "Ligand Receptor Pair" is formed when two macromolecules have combined through molecular recognition to form a complex.

Other examples of receptors which can be investigated by this invention include but are not restricted to:

a) Microorganism receptors: Determination of ligands which bind to receptors, such as specific transport proteins or enzymes essential to survival of microorganisms, is useful in developing a new class of antibiotics. Of particular value would be antibiotics against opportunistic fungi, protozoa, and those bacteria resistant to the antibiotics in current use.

b) Enzymes: For instance, one type of receptor includes the binding site of enzymes such as the enzymes responsible for cleaving neurotransmitters; determination of ligands which bind to certain receptors to modulate the action of the enzymes which cleave the different neurotransmitters is useful in the development of drugs which can be used in the treatment of disorders of neurotransmission.

c) Antibodies: For instance, the invention may be useful in investigating the ligand-binding site on the antibody molecule which combines with the epitope of an antigen of interest;

determining a sequence that mimics an antigenic epitope may lead to the development of vaccines of which the immunogen is based on one or more of such sequences or lead to the development of related diagnostic agents or compounds useful in therapeutic treatments such as for auto-immune diseases (e.g., by blocking the binding of the "self" antibodies).

d) Nucleic Acids: Sequences of nucleic acids may be synthesized to establish DNA or RNA binding sequences.

e) Catalytic Polypeptides: Polymers, preferably polypeptides, which are capable of promoting a chemical reaction involving the conversion of one or more reactants to one or more products. Such polypeptides generally include a binding site specific for at least one reactant or reaction intermediate and an active functionality proximate to the binding site, which functionality is capable of chemically modifying the bound reactant. Catalytic polypeptides are described in, for example, U.S. ~~application Serial No. 404,930~~ Patent No. 5,215,899, which is incorporated herein by reference for all purposes.

f) Hormone receptors: ~~For instance~~, the receptors for insulin and growth hormone. Determination of the ligands which bind with high affinity to a receptor is useful in the development of, for example, an oral replacement of the daily injections which diabetics must take to relieve the symptoms of diabetes, and in the other case, a replacement for the scarce human growth hormone which can only be obtained from cadavers or by recombinant DNA technology. Other examples are the vasoconstrictive hormone receptors; determination of those ligands which

bind to a receptor may lead to the development of drugs to control blood pressure.

g) Opiate receptors: Determination of ligands which bind to the opiate receptors in the brain is useful in the development of less-addictive replacements for morphine and related drugs.

8. Substrate: A material having a rigid or semi-rigid surface. In many embodiments, at least one surface of the substrate will be substantially flat, although in some embodiments it may be desirable to physically separate synthesis regions for different polymers with, for example, wells, raised regions, etched trenches, or the like. According to other embodiments, small beads may be provided on the surface which may be released upon completion of the synthesis.

9. Protective Group: A material which is chemically bound to a monomer unit and which may be removed upon selective exposure to an activator such as electromagnetic radiation. Examples of protective groups with utility herein include those comprising nitropiperonyl, pyrenylmethoxy-carbonyl, nitroveratryl, nitrobenzyl, dimethyl dimethoxybenzyl, 5-bromo-7-nitroindolinyl, o-hydroxy-α-methyl cinnamoyl, and 2-oxymethylene anthraquinone.

10. Predefined Region: A predefined region is a localized area on a surface which is, was, or is intended to be activated for formation of a polymer. The predefined region may have any convenient shape, e.g., circular, rectangular, elliptical, wedge-shaped, etc. For the sake of brevity herein, "predefined regions" are sometimes referred to simply as "regions."

11. Substantially Pure: A polymer is considered to be "substantially pure" within a predefined region of a substrate when it exhibits characteristics that distinguish it from other predefined regions. Typically, purity will be measured in terms of biological activity or function as a result of uniform sequence. Such characteristics will typically be measured by way of binding with a selected ligand or receptor.

12. Activator refers to an energy source adapted to render a group active and which is directed from a source to a predefined location on a substrate. A primary illustration of an activator is light. Other examples of activators include ion beams, electric fields, magnetic fields, electron beams, x-ray, and the like.

13. Binary Synthesis Strategy refers to an ordered strategy for parallel synthesis of diverse polymer sequences by sequential addition of reagents which may be represented by a reactant matrix, and a switch matrix, the product of which is a product matrix. A reactant matrix is a 1 x n matrix of the building blocks to be added. The switch matrix ~~is all or a subset of the binary numbers, preferably ordered, between 1 and n arranged in columns~~ elements of the are 1 n. In preferred embodiments, a binary strategy is one in which at least two successive steps illuminate half of a region of interest on the substrate. In most preferred embodiments, binary synthesis refers to a synthesis strategy which also factors a previous addition step. For example, a strategy in which a switch matrix for a masking strategy halves regions that were previously illuminated, illuminating about half of the previously illuminated region and protecting the remaining half (while also protecting

about half of previously protected regions and illuminating about half of previously protected regions). It will be recognized that binary rounds may be interspersed with non-binary rounds and that only a portion of a substrate may be subjected to a binary scheme, but will still be considered to be a binary masking scheme within the definition herein. A binary "masking" strategy is a binary synthesis which uses light to remove protective groups from materials for addition of other materials such as amino acids. In preferred embodiments, selected columns of the switch matrix are arranged in order of increasing binary numbers in the columns of the switch matrix.

14. Linker refers to a molecule or group of molecules attached to a substrate and spacing a synthesized polymer from the substrate for exposure/binding to a receptor.

II. General

The present invention provides synthetic strategies and devices for the creation of large scale chemical diversity. Solid-phase chemistry, photolabile protecting groups, and photolithography are brought together to achieve light-directed spatially-addressable parallel chemical synthesis in preferred embodiments.

The invention is described herein for purposes of illustration primarily with regard to the preparation of peptides and nucleotides, but could readily be applied in the preparation of other polymers. Such polymers include, for example, both linear and cyclic polymers of nucleic acids, polysaccharides, phospholipids, and peptides having either $\alpha$-, $\beta$-, or $\omega$-amino acids, heteropolymers in which a known drug is covalently bound to any of the above, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene

sulfides, polysiloxanes, polyimides, polyacetates, or other polymers which will be apparent upon review of this disclosure. It will be recognized further, that illustrations herein are primarily with reference to C- to N-terminal synthesis, but the invention could readily be applied to N- to C-terminal synthesis without departing from the scope of the invention.

A. Deprotection and Addition

The present invention uses a masked light source or other activator to direct the simultaneous synthesis of many different chemical compounds. Fig. 1 is a flow chart illustrating the process of forming chemical compounds according to one embodiment of the invention. Synthesis occurs on a solid support 2. A pattern of illumination through a mask 4a using a light source 6 determines which regions of the support are activated for chemical coupling. In one preferred embodiment activation is accomplished by using light to remove photolabile protecting groups from selected areas of the substrate.

After deprotection, a first of a set of building blocks (indicated by "A" in Fig. 1), each bearing a photolabile protecting group (indicated by "X") is exposed to the surface of the substrate and it reacts with regions that were addressed by light in the preceding step. The substrate is then illuminated through a second mask 4b, which activates another region for reaction with a second protected building block "B". The pattern of masks used in these illuminations and the sequence of reactants define the ultimate products and their locations, resulting in diverse sequences at predefined locations, as shown with the sequences ACEG and BDFH in the lower portion of Fig. 1. Preferred embodiments of the invention take advantage of combinatorial masking strategies to form a large number of compounds in a small number of chemical steps.

A high degree of miniaturization is possible because the density of compounds is determined largely with regard to spatial addressability of the activator, in one case the diffraction of light. Each compound is physically accessible and its position is precisely known. Hence, the array is spatially-addressable and its interactions with other molecules can be assessed.

In a particular embodiment shown in Fig. 1, the substrate contains amino groups that are blocked with a photolabile protecting group. Amino acid sequences are made accessible for coupling to a receptor by removal of the photoprotective groups.

When a polymer sequence to be synthesized is, for example, a polypeptide, amino groups at the ends of linkers attached to a glass substrate are derivatized with nitroveratryloxycarbonyl (NVOC), a photoremovable protecting group. The linker molecules may be, for example, aryl acetylene, ethylene glycol oligomers containing from 2-10 monomers, diamines, diacids, amino acids, or combinations thereof. Photodeprotection is effected by illumination of the substrate through, for example, a mask wherein the pattern has transparent regions with dimensions of, for example, less than 1 $cm^2$, $10^{-1}$ $cm^2$, $10^{-2}$ $cm^2$, $10^{-3}$ $cm^2$, $10^{-4}$ $cm^2$, $10^{-5}$ $cm^2$, $10^{-6}$ $cm^2$, $10^{-7}$ $cm^2$, $10^{-8}$ $cm^2$, or $10^{-10}$ $cm^2$. In a preferred embodiment, the regions are between about 10x10 $\mu m$ and 500x500 $\mu m$. According to some embodiments, the masks are arranged to produce a checkerboard array of polymers, although any one of a variety of geometric configurations may be utilized.

1. Example

In one example of the invention, free amino groups were fluorescently labelled by treatment of the entire substrate surface with fluorescein isothiocynate (FITC) after photodeprotection. Glass microscope slides were cleaned, aminated by treatment with

0.1% aminopropyltriethoxysilane in 95% ethanol, and incubated at 110°C for 20 min. The aminated surface of the slide was then exposed to a 30 mM solution of the N-hydroxysuccinimide ester of NVOC-GABA (nitroveratryloxycarbonyl-r-amino butyric acid) in DMF. The NVOC protecting group was photolytically removed by imaging the 365 nm output from a Hg arc lamp through a chrome on glass 100 μm checkerboard mask onto the substrate for 20 min at a power density of 12 mW/cm$^2$. The exposed surface was then treated with 1 mM FITC in DMF. The substrate surface was scanned in an epi-fluorescence microscope (Zeiss Axioskop 20) using 488 nm excitation from an argon ion laser (Spectra-Physics model 2025). The fluorescence emission above 520 nm was detected by a cooled photomultiplier (Hamamatsu 943-02) operated in a photon counting mode. Fluorescence intensity was translated into a color display with red in the highest intensity and black in the lowest intensity areas. The presence of a high-contrast fluorescent checkerboard pattern of 100x100 μm elements revealed that free amino groups were generated in specific regions by spatially-localized photodeprotection.

2. Example

Fig. 2 is a flow chart illustrating another example of the invention. Carboxy-activated NVOC-leucine was allowed to react with an aminated substrate. The carboxy activated HOBT ester of leucine and other amino acids used in this synthesis was formed by mixing 0.25 mmol of the NVOC amino protected amino acid with 37 mg HOBT (1-hydroxybenzotriazole), 111 mg BOP (benzotriazolyl-n-oxy-tris (dimethylamino)-phosphoniumhexa-fluorophosphate) and 86 μl DIEA (diisopropylethylamine) in 2.5 ml DMF. The NVOC protecting group was removed by uniform illumination. Carboxy-activated NVOC-phenylalanine was coupled to the exposed amino groups for 2 hours at room temperature,

and then washed with DMF and methylene chloride. Two unmasked cycles of photodeprotection and coupling with carboxy-activated NVOC-glycine were carried out. The surface was then illuminated through a chrome on glass 50 μm checkerboard pattern mask. Carboxy-activated Nα-tBOC-O-tButyl-L-tyrosine was then added. The entire surface was uniformly illuminated to photolyze the remaining NVOC groups. Finally, carboxy-activated NVOC-L-proline was added, the NVOC group was removed by illumination, and the t-BOC and t-butyl protecting groups were removed with TFA. After removal of the protecting groups, the surface consisted of a 50 μm checkerboard array of Tyr-Gly-Gly-Phe-Leu (YGGFL) (SEQ ID NO:1) and Pro-Gly-Gly-Phe-Leu (PGGFL) (SEQ ID NO:2).

B. Antibody Recognition

In one preferred embodiment the substrate is used to determine which of a plurality of amino acid sequences is recognized by an antibody of interest.

1. Example

In one example, the array of pentapeptides in the example illustrated in Fig. 2 was probed with a mouse monoclonal antibody directed against β-endorphin. This antibody (called 3E7) is known to bind YGGFL and YGGFM (SEQ ID NO:4) with nanomolar affinity and is discussed in Meo et al., Proc. Natl. Acad. Sci. USA (1983) 80:4084, which is incorporated by reference herein for all purposes. This antibody requires the amino terminal tyrosine for high affinity binding. The array of peptides formed as described in Fig. 2 was incubated with a 2 μg/ml mouse monoclonal antibody (3E7) known to recognize YGGFL. 3E7 does not bind PGGFL. A second incubation with fluoresceinated goat anti-mouse antibody labeled the regions that bound 3E7. The surface was scanned with an epi-fluorescence microscope. The results showed alternating bright and dark 50 μm squares indicating that

YGGFL and PGGFL were synthesized in geometric array determined by the mask. A high contrast (>12:1 intensity ratio) fluorescence checkerboard image shows that (a) YGGFL and PGGFL were synthesized in alternate 50 μm squares, (b) YGGFL attached to the surface is accessible for binding to antibody 3E7, and (c) antibody 3E7 does not bind to PGGFL.

A three-dimensional representation of the fluorescence intensity data in a portion of the checkboard is shown in Fig. 3. This figure shows that the border between synthesis sites is sharp. The height of each spike in this display is linearly proportional to the integrated fluorescence intensity in a 2.5 μm pixel. The transition between PGGFL and YGGFL occurs within two spikes (5 μm). There is little variation in the fluorescence intensity of different YGGFL squares. The mean intensity of sixteen YGGFL synthesis sites was $2.03 \times 10^5$ counts and the standard deviation was $9.6 \times 10^3$ counts.

III. Synthesis

A. Reactor System

Fig. 4 schematically illustrates a device used to synthesize diverse polymer sequences on a substrate. The device includes an automated peptide synthesizer 401. The automated peptide synthesizer is a device which flows selected reagents through a flow cell 402 under the direction of a computer 404. In a preferred embodiment the synthesizer is an ABI Peptide Synthesizer, model no. 431A. The computer may be selected from a wide variety of computers or discrete logic including for, example, an IBM PC-AT or similar computer linked with appropriate internal control systems in the peptide synthesizer. The PC is provided with signals from the board computer indicative of, for example, the end of a coupling cycle.

Substrate 406 is mounted on the flow cell, forming a cavity between the substrate and the flow cell. Selected reagents flow through this cavity from the peptide synthesizer at selected times, forming an array of peptides on the face of the substrate in the cavity. Mounted above the substrate, and preferably in contact with the substrate is a mask 408. Mask 408 is transparent in selected regions to a selected wavelength of light and is opaque in other regions to the selected wavelength of light. The mask is illuminated with a light source 410 such as a UV light source. In one specific embodiment the light source 410 is a model no. 82420 made by Oriel. The mask is held and translated by an x-y-z translation stage 412 such as an x-y translation stage made by Newport Corp. The computer coordinates action of the peptide synthesizer, x-y translation stage, and light source. Of course, the invention may be used in some embodiments with translation of the substrate instead of the mask.

In operation, the substrate is mounted on the reactor cavity. The slide, with its surface protected by a suitable photo removable protective group, is exposed to light at selected locations by positioning the mask and illuminating the light source for a desired period of time (such as, for example, 1 sec to 60 min in the case of peptide synthesis). A selected peptide or other monomer/polymer is pumped through the reactor cavity by the peptide synthesizer for binding at the selected locations on the substrate. After a selected reaction time (such as about 1 sec to 300 min in the case of peptide reactions) the monomer is washed from the system, the mask is appropriately repositioned or replaced, and the cycle is repeated. In most embodiments of the invention, reactions may be conducted at or near ambient temperature.

Figs 5a and 5b are flow charts of the software used in operation of the reactor system. At step 502 the peptide synthesis software is initialized. At step 504 the system calibrates positioners on the x-y translation stage and begins a main loop. At step 506 the system determines which, if any, of the function keys on the computer have been pressed. If F1 has been pressed, the system prompts the user for input of a desired synthesis process. If the user enters F2, the system allows a user to edit a file for a synthesis process at step 510. If the user enters F3 the system loads a process from a disk at step 512. If the user enters F4 the system saves an entered or edited process to disk at step 514. If the user selects F5 the current process is displayed at step 516 while selection of F6 starts the main portion of the program, i.e., the actual synthesis according to the selected process. If the user selects F7 the system displays the location of the synthesized peptides, while pressing F10 returns the user to the disk operating system.

Fig. 5b illustrates the synthesis step 518 in greater detail. The main loop of the program is started in which the system first moves the mask to a next position at step 526. During the main loop of the program, necessary chemicals flow through the reaction cell under the direction of the on-board computer in the peptide synthesizer. At step 528 the system then waits for an exposure command and, upon receipt of the exposure command exposes the substrate for a desired time at step 530. When an acknowledge of exposure complete is received at step 532 the system determines if the process is complete at step 534 and, if so, waits for additional keyboard input at step 536 and, thereafter, exits the perform synthesis process.

A computer program used for operation of the system described above is included as microfiche Appendix A (Copyright, 1990, Affymax Technologies N.V.,

all rights reserved). The program is written in Turbo C++ (Borland Int'l) and has been implemented in an IBM compatible system. The motor control software is adapted from software produced by Newport Corporation. It will be recognized that a large variety of programming languages could be utilized without departing from the scope of the invention herein. Certain calls are made to a graphics program in "Programmer Guide to PC and PS2 Video Systems" (Wilton, Microsoft Press, 1987), which is incorporated herein by reference for all purposes.

Alignment of the mask is achieved by one of two methods in preferred embodiments. In a first embodiment the system relies upon relative alignment of the various components, which is normally acceptable since x-y-z translation stages are capable of sufficient accuracy for the purposes herein. In alternative embodiments, alignment marks on the substrate are coupled to a CCD device for appropriate alignment.

According to some embodiments, pure reagents are not added at each step, or complete photolysis of the protective groups is not provided at each step. According to these embodiments, multiple products will be formed in each synthesis site. For example, if the monomers A and B are mixed during a synthesis step, A and B will bind to deprotected regions, roughly in proportion to their concentration in solution. Hence, a mixture of compounds will be formed in a synthesis region. A substrate formed with mixtures of compounds in various synthesis regions may be used to perform, for example, an initial screening of a large number of compounds, after which a smaller number of compounds in regions which exhibit high binding affinity are further screened. Similar results may be obtained by only partially photylizing a region, adding a first monomer, re-photylizing the same region, and exposing the region to a second monomer.

27

B. Binary Synthesis Strategy

In a light-directed chemical synthesis, the products formed depend on the pattern and order of masks, and on the order of reactants. To make a set of products there will in general be "n" possible masking schemes. In preferred embodiments of the invention herein a binary synthesis strategy is utilized. The binary synthesis strategy is illustrated herein primarily with regard to a masking strategy, although it will be applicable to other polymer synthesis strategies such as the pin strategy, and the like.

In a binary synthesis strategy, the substrate is irradiated with a first mask, exposed to a first building block, irradiated with a second mask, exposed to a second building block, etc. Each combination of masked irradiation and exposure to a building block is referred to herein as a "cycle."

In a preferred binary masking scheme, the masks for each cycle allow irradiation of half of a region of interest on the substrate and protection of the remaining half of the region of interest. By "half" it is intended herein not to mean exactly one-half the region of interest, but instead a large fraction of the region of interest such as from about 30 to 70 percent of the region of interest. It will be understood that the entire masking scheme need not take a binary form; instead non-binary cycles may be introduced as desired between binary cycles.

In preferred embodiments of the binary masking scheme, a given cycle illuminates only about half of the region which was illuminated in a previous cycle, while protecting the remaining half of the illuminated portion from the previous cycle. Conversely, in such preferred embodiments, a given cycle illuminates half of the region which was protected in the previous cycle and protects half the region which was protected in a previous cycle.

72

In the synthesis strategy disclosed herein, the longest length ($\ell$) of the synthesized polymers is $\ell = n/a$; where n is the number of cycles and a is the number of chemical building blocks (note that a given building block may be repeated). The total number of possible compounds synthesized (k) will be $k = a^{\ell}$. For example, if a = 20 and $\ell$ = 5, n = 100 and $k = 3.2 \times 10^6$.

The synthesis strategy is most readily illustrated and handled in matrix notation. At each synthesis site, the determination of whether to add a given monomer is a binary process. Therefore, each product element $P_j$ is given by the dot product of two vectors, a chemical reactant vector, e.g., C = [A,B,C,D], and a binary vector $\sigma_j$. Inspection of the products in the example below for a four-step synthesis, shows that in one four-step synthesis $\sigma_1$ = [1,0,1,0], $\sigma_2$ = [1,0,0,1], $\sigma_3$ = [0,1,1,0], and $\sigma_4$ = [0,1,0,1], where a 1 indicates illumination and a 0 indicates protection. Therefore, it becomes possible to build a "switch matrix" S from the column vectors $\sigma_j$ (j = 1,k where k is the number of products).

$$S = \begin{matrix} \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{matrix}$$

The outcome P of a synthesis is simply P = CS, the product of the chemical reactant matrix and the switch matrix.

The switch matrix for an n-cycle synthesis yielding k products has n rows and k columns. An important attribute of S is that each row specifies a mask. A two-dimensional mask $m_j$ for the jth chemical step of a synthesis is obtained directly from the jth row of S by placing the elements $s_{j1},\ldots s_{jk}$ into, for example, a

square format. The particular arrangement below provides a square format, although linear or other arrangements may be utilized.

$$S = \begin{matrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{21} & s_{22} & s_{23} & s_{24} \\ s_{31} & s_{32} & s_{33} & s_{34} \\ s_{41} & s_{42} & s_{43} & s_{44} \end{matrix} \qquad m_j = \begin{matrix} s_{j1} & s_{j2} \\ s_{j3} & s_{j4} \end{matrix}$$

Of course, compounds formed in a light-activated synthesis can be positioned in any defined geometric array. A square or rectangular matrix is convenient but not required. The rows of the switch matrix may be transformed into any convenient array as long as equivalent transformations are used for each row.

For example, the masks in the four-step synthesis below are then denoted by:

$$m_1 = \begin{matrix} 1 & 1 \\ 0 & 0 \end{matrix} \qquad m_2 = \begin{matrix} 0 & 0 \\ 1 & 1 \end{matrix} \qquad m_3 = \begin{matrix} 1 & 0 \\ 1 & 0 \end{matrix} \qquad m_4 = \begin{matrix} 0 & 1 \\ 0 & 1 \end{matrix}$$

where 1 denotes illumination (activation) and 0 denotes no illumination.

The matrix representation is used to generate a desired set of products and product maps in preferred embodiments. Each compound is defined by the product of the chemical vector and a particular switch vector. Therefore, for each synthesis address, one simply saves the switch vector, assembles all of them into a switch matrix, and extracts each of the rows to form the masks.

In some cases, particular product distributions or a maximal number of products are desired. For example, for C = [A,B,C,D], any switch vector ($\sigma_j$) consists of four bits. Sixteen four-bit vectors exist. Hence, a maximum of 16 different products can be made by sequential addition of the reagents [A,B,C,D]. These 16 column vectors can be assembled in 16! different ways to

form a switch matrix. The order of the column vectors defines the masking patterns, and therefore, the spatial ordering of products but not their makeup. One ordering of these columns gives the following switch matrix (in which "null" (Ø) additions are included in brackets for the sake of completeness, although such null additions are elsewhere ignored herein):

$$S = \begin{matrix} \sigma 1 & & & & & & & & & & & & & & & \sigma_{16} & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & A \\ [0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1] & Ø \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & B \\ [0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1] & Ø \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & C \\ [0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1] & Ø \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & D \\ [0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1] & Ø \end{matrix}$$

The columns of S according to this aspect of the invention are the binary representations of the numbers 15 to 0. The sixteen products of this binary synthesis are ABCD, ABC, ABD, AB, ACD, AC, AD, A, BCD, BC, BD, B, CD, C, D, and Ø (null). Also note that each of the switch vectors from the four-step synthesis masks above (and hence the synthesis products) are present in the four bit binary switch matrix. (See columns 6, 7, 10, and 11)

This synthesis procedure provides an easy way for mapping the completed products. The products in the various locations on the substrate are simply defined by the columns of the switch matrix (the first column indicating, for example, that the product ABCD will be present in the upper left-hand location of the substrate). Furthermore, if only selected desired products are to be made, the mask sequence can be derived by extracting the columns with the desired sequences. For example, to form the product set ABCD, ABD, ACD, AD,

BCD, BD, CD, and D, the masks are formed by use of a switch matrix with only the 1st, 3rd, 5th, 7th, 9th, 11th, 13th, and 15th columns arranged into the switch matrix:

$$S = \begin{matrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{matrix}$$

To form all of the polymers of length 4, the reactant matrix [ABCDABCDABCDABCD] is used. The switch matrix will be formed from a matrix of the binary numbers from 0 to $2^{16}$ arranged in columns. The columns having four monomers are than selected and arranged into a switch matrix. Therefore, it is seen that the binary switch matrix in general will provide a representation of all the products which can be made from an n-step synthesis, from which the desired products are then extracted.

The rows of the binary switch matrix will, in preferred embodiments, have the property that each masking step illuminates half of the synthesis area. Each masking step also factors the preceding masking step; that is, half of the region that was illuminated in the preceding step is again illuminated, whereas the other half is not. Half of the region that was unilluminated in the preceding step is also illuminated, whereas the other half is not. Thus, masking is recursive. The masks are constructed, as described previously, by extracting the elements of each row and placing them in a square array. For example, the four masks in S for a four-step synthesis are:

$$m_1 = \begin{matrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix} \quad m_2 = \begin{matrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{matrix} \quad m_3 = \begin{matrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{matrix} \quad m_4 = \begin{matrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{matrix}$$

The recursive factoring of masks allows the products of a light-directed synthesis to be represented by a polynomial. (Some light activated syntheses can only be denoted by irreducible, i.e., prime polynomials.) For example, the polynomial corresponding to the top synthesis of Fig. 9a (discussed below) is

$$P = (A + B)(C + D)$$

A reaction polynomial may be expanded as though it were an algebraic expression, provided that the order of joining of reactants $X_1$ and $X_2$ is preserved ($X_1X_2 \neq X_2X_1$), i.e., the products are not commutative. The product then is AC + AD + BC + BD. The polynomial explicitly specifies the reactants and implicitly specifies the mask for each step. Each pair of parentheses demarcates a round of synthesis. The chemical reactants of a round (e.g., A and B) react at nonoverlapping sites and hence cannot combine with one other. The synthesis area is divided equally amongst the elements of a round (e.g., A is directed to one-half of the area and B to the other half). Hence, the masks for a round (e.g., the masks $m_A$ and $m_B$) are orthogonal and form an orthonormal set. The polynomial notation also signifies that each element in a round is to be joined to each element of the next round (e.g., A with C, A with D, B with C, and B with D). This is accomplished by having $m_C$ overlap $m_A$ an $m_B$ equally, and likewise for $m_D$. Because C and D are elements of a round, $m_C$ and $m_D$ are orthogonal to each other and form an orthonormal set.

The polynomial representation of the binary synthesis described above, in which 16 products are made from 4 reactants, is

$$P = (A + \emptyset)\ (B + \emptyset)\ (C + \emptyset)\ (D + \emptyset)$$

which gives ABCD, ABC, ABD, AB, ACD, AC, AD, A, BCD, BC, BD, B, CD, C, D, and Ø when expanded (with the rule that ØX = X and XØ = X, and remembering that joining is ordered). In a binary synthesis, each round contains one reactant and one null (denoted by Ø). Half of the synthesis area receives the reactant and the other half receives nothing. Each mask overlaps every other mask equally.

Binary rounds and non-binary rounds can be interspersed as desired, as in

$$P = (A + Ø)\ (B)\ (C + D + Ø)\ (E + F + G)$$

The 18 compounds formed are ABCE, ABCF, ABCG, ABDE, ABDF, ABDG, ABE, ABF, ABG, BCE, BCF, BCG, BDE, BDF, BDG, BE, BF, and BG. The switch matrix S for this 7-step synthesis is

$$S = \begin{matrix} 1&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0 \\ 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 1&1&1&0&0&0&0&0&1&1&1&0&0&0&0&0&0 \\ 0&0&0&1&1&1&0&0&0&0&0&0&1&1&1&0&0 \\ 1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0 \\ 0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1 \\ 0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&1 \end{matrix}$$

The round denoted by (B) places B in all products because the reaction area was uniformly activated (the mask for B consisted entirely of 1's).

The number of compounds k formed in a synthesis consisting of r rounds, in which the ith round has $b_i$ chemical reactants and $z_i$ nulls, is

$$k = \Sigma\ (b_i + z_i)$$

and the number of chemical steps n is

$$n = \Sigma b_i$$

The number of compounds synthesized when b = a and z = 0 in all rounds is $a^{n/a}$, compared with $2^n$ for a binary synthesis. For n = 20 and a = 5, 625 compounds (all tetramers) would be formed, compared with $1.049 \times 10^6$ compounds in a binary synthesis with the same number of chemical steps.

It should also be noted that rounds in a polynomial can be nested, as in

$$(A + (B+\emptyset)(C+\emptyset))\ (D+\emptyset)$$

The products are AD, BCD, BD, CD, D, A, BC, B, C, and Ø.

Binary syntheses are attractive for two reasons. First, they generate the maximal number of products ($2^n$) for a given number of chemical steps (n). For four reactants, 16 compounds are formed in the binary synthesis, whereas only 4 are made when each round has two reactants. A 10-step binary synthesis yields 1,024 compounds, and a 20-step synthesis yields 1,048,576. Second, products formed in a binary synthesis are a complete nested set with lengths ranging from 0 to n. All compounds that can be formed by deleting one or more units from the longest product (the n-mer) are present. Contained within the binary set are the smaller sets that would be formed from the same reactants using any other set of masks (e.g., AC, AD, BC, and BD formed in the synthesis shown in Fig. 6 are present in the set of 16 formed by the binary synthesis). In some cases, however, the experimentally achievable spatial resolution may not suffice to accommodate all the compounds formed. Therefore, practical limitations may require one to select a particular subset of the possible switch vectors for a given synthesis.

1. Example

Fig. 6 illustrates a synthesis with binary masking scheme. The binary masking scheme provides the greatest number of sequences for a given number of cycles. According to this embodiment, a mask m1 allows illumination of half of the substrate. The substrate is then exposed to the building block A, which binds at the illuminated regions.

Thereafter, the mask m2 allows illumination of half of the previously illuminated region, while protecting half of the previously illuminated region. The building block B is then added, which binds at the illuminated regions from m2.

The process continues with masks m3, m4, and m5, resulting in the product array shown in the bottom portion of the figure. The process generates 32 (2 raised to the power of the number of monomers) sequences with 5 (the number of monomers) cycles.

2. Example

Fig. 7 illustrates another preferred binary masking scheme which is referred to herein as the gray code masking scheme. According to this embodiment, the masks m1 to m5 are selected such that a side of any given synthesis region is defined by the edge of only one mask. The site at which the sequence BCDE is formed, for example, has its right edge defined by m5 and its left side formed by mask m4 (and no other mask is aligned on the sides of this site). Accordingly, problems created by misalignment, diffusion of light under the mask and the like will be minimized.

3. Example

Fig. 8 illustrates another binary masking scheme. According to this scheme, referred to herein as a modified gray code masking scheme, the number of masks needed is minimized. For example, the mask m2 could be

the same mask as m1 and simply translated laterally. Similarly, the mask m4 could be the same as mask m3 and simply translated laterally.

4. Example

A four-step synthesis is shown in Fig. 9a. The reactants are the ordered set (A,B,C,D). In the first cycle, illumination through $m_1$ activates the upper half of the synthesis area. Building block A is then added to give the distribution 602. Illumination through mask $m_2$ (which activates the lower half), followed by addition of B yields the next intermediate distribution 604. C is added after illumination through $m_3$ (which activates the left half) giving the distribution 604, and D after illumination through $m_4$ (which activates the right half), to yield the final product pattern 608 (AC,AD,BC,BD).

5. Example

The above masking strategy for the synthesis may be extended for all 400 dipeptides from the 20 naturally occurring amino acids as shown in Fig. 9b. The synthesis consists of two rounds, with 20 photolysis and chemical coupling cycles per round. In the first cycle of round 1, mask 1 activates 1/20th of the substrate for coupling with the first of 20 amino acids. Nineteen subsequent illumination/coupling cycles in round 1 yield a substrate consisting of 20 rectangular stripes each bearing a distinct member of the 20 amino acids. The masks of round 2 are perpendicular to round 1 masks and therefore a single illumination/coupling cycle in round 2 yields 20 dipeptides. The 20 illumination/coupling cycles of round 2 complete the synthesis of the 400 dipeptides.

6. Example

The power of the binary masking strategy can be appreciated by the outcome of a 10-step synthesis that

produced 1,024 peptides. The polynomial expression for this 10-step binary synthesis was:

(f+Ø) (Y+Ø) (G+Ø) (A+Ø) (G+Ø) (T+Ø) (F+Ø) (L+Ø) (S+Ø) (F+Ø)

Each peptide occupied a 400x400 μm square. A 32x32 peptide array (1,024 peptides, including the null peptide and 10 peptides of $\ell$ = 1, and a limited number of duplicates) was clearly evident in a fluorescence scan following side group deprotection and treatment with the antibody 3E7 and fluorescinated antibody. Each synthesis site was a 400x400 μm square.

The scan showed a range of fluorescence intensities, from a background value of 3,300 counts to 22,400 counts in the brightest square (x = 20, y = 9). Only 15 compounds exhibited an intensity greater than 12,300 counts. The median value of the array was 4,800 counts.

The identity of each peptide in the array could be determined from its x and y coordinates (each range from 0 to 31) and the map of Fig. 10. The chemical units at positions 2, 5, 6, 9, and 10 are specified by the y coordinate and those at positions 1, 3, 4, 7, 8 by the x coordinate. All but one of the peptides was shorter than 10 residues. For example, the peptide at x = 12 and y = 3 is YGAGF (SEQ ID NO:2) (positions 1, 6, 8, 9, and 10 are nulls). YGAFLS (SEQ ID NO:4), the brightest element of the array, is at x = 20 and y = 9.

It is often desirable to deduce a binding affinity of a given peptide from the measured fluorescence intensity. Conceptually, the simplest case is one in which a single peptide binds to a univalent antibody molecule. The fluorescence scan is carried out after the slide is washed with buffer for a defined time. The order of fluorescence intensities is then a measure primarily of the relative dissociation rates of the antibody-peptide complexes. If the on-rate constants are

the same (e.g., if they are diffusion-controlled), the order of fluorescence intensities will correspond to the order of binding affinities. However, the situation is sometimes more complex because a bivalent primary antibody and a bivalent secondary antibody are used. The density of peptides in a synthesis area corresponded to a mean separation of ~7 nm, which would allow multivalent antibody-peptide interactions. Hence, fluorescence intensities obtained according to the method herein will often be a qualitative indicator of binding affinity.

Another important consideration is the fidelity of synthesis. Deletions are produced by incomplete photodeprotection or incomplete coupling. The coupling yield per cycle in these experiments is typically between 85% and 95%. Implementing the switch matrix by masking is imperfect because of light diffraction, internal reflection, and scattering. Consequently, stowaways (chemical units that should not be on board) arise by unintended illumination of regions that should be dark. A binary synthesis array contains many of the controls needed to assess the fidelity of a synthesis. For example, the fluorescence signal from a synthesis area nominally containing a tetrapeptide ABCD could come from a tripeptide deletion impurity such as ACD. Such an artifact would be ruled out by the finding that the fluorescence intensity of the ACD site is less than that of the ABCD site.

The fifteen most highly labelled peptides in the array obtained with the synthesis of 1,024 peptides described above, were YGAFLS, YGAFS, YGAFL, YGGFLS, YGAF, YGALS, YGGFS, YGAL, YGAFLF, YGAF, YGAFF, YGGLS, YGGFL, YGAFSF, and YGAFLSF. A striking feature is that all fifteen begin with YG, which agrees with previous work showing that an amino-terminal tyrosine is a key determinant of binding. Residue 3 of this set is either A or G, and residue 4 is either F or L. The exclusion of S and T from these positions is clear cut. The finding

that the preferred sequence is YG (A/G) (F/L) fits nicely with the outcome of a study in which a very large library of peptides on phage generated by recombinant DNA methods was screened for binding to antibody 3E7 (see Cwirla et al., Proc. Natl. Acad. Sci. USA, (1990) 87:6378, incorporated herein by reference). Additional binary syntheses based on leads from peptides on phage experiments show that YGAFMQ, YGAFM, and YGAFQ give stronger fluorescence signals than does YGGFM, the immunogen used to obtain antibody 3E7.

Variations on the above masking strategy will be valuable in certain circumstances. For example, if a "kernel" sequence of interest consists of PQR separated from XYZ and that the aim is to synthesize peptides in which these units are separated by a variable number of different residues, the kernel can be placed in each peptide by using a mask that has 1's everywhere. The polynomial representation of a suitable synthesis is:

$$(P)(Q)(R)(A+\emptyset)(B+\emptyset)(C+\emptyset)(D+\emptyset)(X)(Y)(Z)$$

Sixteen peptides will be formed, ranging in length from the 6-mer PQRXYZ to the 10-mer PQRABCDXYZ.

Several other masking strategies will also find value in selected circumstances. By using a particular mask more than once, two or more reactants will appear in the same set of products. For example, suppose that the mask for an 8-step synthesis is

| | |
|---|---|
| A | 11110000 |
| B | 00001111 |
| C | 11001100 |
| D | 00110011 |
| E | 10101010 |
| F | 01010101 |
| G | 11110000 |
| H | 00001111 |

The products are ACEG, ACFG, ADEG, ADFG, BCEH, BCFH, BDEH, and BDFH. A and G always appear together because their additions were directed by the same mask, and likewise for B and H.

C. Linker Selection

According to preferred embodiments the linker molecules used as an intermediary between the synthesized polymers and the substrate are selected for optimum length and/or type for improved binding interaction with a receptor. According to this aspect of the invention diverse linkers of varying length and/or type are synthesized for subsequent attachment of a ligand. Through variations in the length and type of linker, it becomes possible to optimize the binding interaction between an immobilized ligand and its receptor.

The degree of binding between a ligand (peptide, inhibitor, hapten, drug, etc.) and its receptor (enzyme, antibody, etc.) when one of the partners is immobilized on to a substrate will in some embodiments depend on the accessibility of the receptor in solution to the immobilized ligand. The accessibility in turn will depend on the length and/or type of linker molecule employed to immobilize one of the partners. Preferred embodiments of the invention therefore employ the VLSIPS™ technology described herein to generate an array of, preferably, inactive or inert linkers of varying length and/or type, using photochemical protecting groups to selectively expose different regions of the substrate and to build upon chemically-active groups.

In the simplest embodiment of this concept, the same unit is attached to the substrate in varying multiples or lengths in known locations on the substrate via VLSIPS™ techniques to generate an array of polymers of varying length. A single ligand (peptide, drug, hapten, etc.) is attached to each of them, and an assay is performed with the binding site to evaluate the degree of

binding with a receptor that is known to bind to the ligand. In cases where the linker length impacts the ability of the receptor to bind to the ligand, varying levels of binding will be observed. In general, the linker which provides the highest binding will then be used to assay other ligands synthesized in accordance with the techniques herein.

According to other embodiments the binding between a single ligand/receptor pair is evaluated for linkers of diverse monomer sequence. According to these embodiments, the linkers are synthesized in an array in accordance with the techniques herein and have different monomer sequence (and, optionally, different lengths). Thereafter, all of the linker molecules are provided with a ligand known to have at least some binding affinity for a given receptor. The given receptor is then exposed to the ligand and binding affinity is deduced. Linker molecules which provide adequate binding between the ligand and receptor are then utilized in screening studies.

D. Protecting Groups

As discussed above, selectively removable protecting groups allow creation of well defined areas of substrate surface having differing reactivities. Preferably, the protecting groups are selectively removed from the surface by applying a specific activator, such as electromagnetic radiation of a specific wavelength and intensity. More preferably, the specific activator exposes selected areas of surface to remove the protecting groups in the exposed areas.

Protecting groups of the present invention are used in conjunction with solid phase oligomer syntheses, such as peptide syntheses using natural or unnatural amino acids, nucleotide syntheses using deoxyribonucleic and ribonucleic acids, oligosaccharide syntheses, and the like. In addition to protecting the substrate surface